Il corpus parallelo italiano-russo del NKRJa. Progetto di ampliamento, applicazioni e sviluppi

Francesca Biagini, Anna Bonola, Valentina Noseda[1*]

1. *La linguistica dei corpora nel mondo slavo e i corpora paralleli*

La linguistica dei *corpora*, sviluppatasi in ambito anglosassone ormai oltre cinquant'anni fa[2], ha raggiunto il mondo slavo relativamente tardi. I primi *corpora* delle lingue slave sono comparsi tra la fine degli anni Novanta e gli inizi del Duemila. I maggiori *corpora* al momento disponibili per le lingue slave sono i *Corpora* Nazionali delle lingue ceca (CNC)[3], bulgara[4], polacca[5], slovacca[6] e russa[7] (*Nacional'nyj Korpus Russkogo Jazyka* – NKRJa)[8] (Reznikova 2009). Tra questi, per dimensioni e livello di specializzazione, emergono in particolare il CNC e il NKRJa.

Il CNC è un insieme di *corpora*, raccolti sul sito dell'*Institute of Czech National Corpus* grazie al lavoro della Facoltà delle Arti dell'Università Carolina di Praga[9]. Il progetto, avviato nel 1994 (accessibile dal 2000), conta ora più di un miliardo di parole ed è annotato morfologicamente[10]. Il NKRJa, invece, nato

---

[1] * Il contenuto del presente lavoro è da attribuire ai singoli autori come segue: V. Noseda (parr. 1-2), F. Biagini (par. 3), A. Bonola (par. 4).

[2] I primi corpora per la lingua inglese nascono negli anni '60: il *Brown Corpus* negli Stati Uniti e il *London-Lund Corpus* in UK (McEnery, Wilson 2001: 22-23).

[3] <http://www.korpus.cz/> (ultimo accesso: 15.03.19).

[4] <http://search.dcl.bas.bg/> (ultimo accesso: 15.03.19).

[5] <http://nkjp.pl/> (ultimo accesso: 15.03.19).

[6] <http://korpus.juls.savba.sk/> (ultimo accesso: 15.03.19).

[7] <http://rusorpora.ru/> (ultimo accesso: 15.03.19).

[8] Reznikova (2009) cita anche alcuni *corpora* minori: FIDA e FidaPLUS, *Nova beseda* e KoRP per lo sloveno, il *Corpus* nazionale della lingua croata, e quelli della lingua bosniaca e serba.

[9] <https://ucnk.ff.cuni.cz/english/struktura.php> (ultimo accesso: 15.03.19).

[10] Il CNC comprende ora quindici *corpora* sincronici di lingua scritta (alcuni generali e di vasta portata, altri più specifici e ridotti, quali ad esempio l'*ORWELL Corpus*, che include esclusivamente il romanzo *1984* di G. Orwell tradotto in lingua ceca e conta 80.000 parole); sette *corpora* sincronici di lingua parlata, di media entità (da 215.000 a 2.790.000 parole); un *corpus* diacronico composto da poco meno di due milioni di parole e da testi in lingua ceca risalenti a sette secoli diversi; otto *corpora* in altre lingue,

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Francesca Biagini, Anna Bonola, Valentina Noseda, *Il corpus parallelo italiano-russo del NKRJa. Progetto di ampliamento, applicazioni e sviluppi*, pp. 35-45, DOI 10.36253/978-88-6453-910-2.06

nel 2004 (Sičinava 2005) comprende oggi più di cinquecento milioni di parole suddivise tra un *corpus* principale generale (*osnovnoj korpus*) e dieci *sottocorpora* specialistici: *corpus* sintattico, giornalistico, dialettale, didattico, parallelo, poetico, della lingua parlata, della storia dell'accento russo, multimediale, storico. L'intero *corpus* presenta un'annotazione morfo-sintattica, semantica e metatestuale, e dispone di un motore di ricerca altamente avanzato.

Nonostante la crescente diffusione della linguistica dei *corpora*, rimangono ancora relativamente scarse le risorse elettroniche multilinguistiche. Infatti, tra i *corpora* delle lingue slave, solo il CNC e il NKRJa dedicano una sezione ai *corpora* paralleli (d'ora in avanti: CP)[11]. Il CNC comprende il progetto InterCorp[12] con una sezione di *corpora* compilati tradizionalmente (i testi sono selezionati secondo criteri predefiniti) definita *core corpus* (195 milioni di parole), e un insieme di risorse definibili come 'collezioni'[13], tratte da diverse risorse in parte accessibili anche attraverso altri portali e non sottoposte a correzione manuale (es.: Opus[14] o Sketch Engine[15]).

I quindici CP del NKRJa variano tra loro per entità e tipo[16]: il più vasto (russo-inglese) conta quasi 25 milioni di parole, il più piccolo (russo-cinese) ne comprende appena 55.424. Fino al 2015, tra gli ultimi posti si trovava anche il CP italiano-russo (CP ita-ru), nato come *corpus* pilota bi-direzionale comprendente in tutto cinque testi per un totale di circa 700 mila parole.

Considerando le potenzialità offerte dai CP per la ricerca scientifica, e consapevoli del fatto che il *corpus* pilota ita-ru non possedeva, per le sue dimensioni, i requisiti di uno strumento scientificamente valido, abbiamo elaborato un progetto per il suo ampliamento, la cui fase iniziale è stata portata a termine nel dicembre 2015 nell'ambito del lavoro di dottorato descritto in Noseda 2016.

2. *L'ampliamento del CP italiano-russo nel NKRJA*

L'ampliamento del CP ita-ru è stato preceduto da un'analisi comparata dei migliori CP esistenti, tra cui si è evidenziato, per struttura e rappresentatività,

---

tra cui l'italiano (ItWac). Si tratta soprattutto di *web corpora*, ricavati da testi estratti automaticamente dalla rete (Baroni *et al*. 2009).

[11] Selezioni di testi in una data lingua (A) allineati – generalmente a livello di frase – con la rispettiva traduzione in lingua B.

[12] Accessibile dal sito: <https://kontext.korpus.cz/first_form> (ultimo accesso: 15.03.19).

[13] Secondo la definizione di EAGLES (*Expert Advisory Group on Language Engineering Standards*): "words such as 'collection' and 'archive' refer to sets of texts that do not need to be selected, or do not need to be ordered, or the selection and/or ordering do not need to be on linguistic criteria. They are therefore quite unlike corpora" (<http://www.ilc.cnr.it/EAGLES96/intro.html>, ultimo accesso: 15.03.19).

[14] <http://opus.lingfil.uu.se/> (ultimo accesso: 15.03.19).

[15] <https://www.sketchengine.co.uk/> (ultimo accesso: 15.03.19).

[16] Alcuni contengono solo testi letterari, altri hanno testi di diversi generi, oltre a variare i criteri di selezione dei testi.

l'*English-Norwegian Parallel Corpus* (ENPC), che è servito anche da modello per progettare un CP ita-ru che sia:

1. *Corpus* di monitoraggio (e non statico): il *corpus* verrà gradualmente arricchito di nuovi testi.
2. *Corpus* bi-direzionale: comprende originali russi con traduzione italiana e viceversa, così da costituire allo stesso tempo un *corpus* parallelo e comparabile, in cui sia possibile confrontare non solo originali in lingua A e traduzioni in lingua B, ma anche testi originali e traduzioni in una stessa lingua.
3. *Corpus* diacronico: con testi di periodi diversi (XIX, XX e XXI secolo).
4. *Corpus* di estratti (di 10.000-15.000 parole): la scelta di inserire solo alcuni brani, anziché opere intere era inizialmente legata all'ingente lavoro richiesto dall'allineamento e a problemi di *copyright*; tuttavia, in seguito, testando il CP con alcune analisi contrastive (Noseda 2016), ci si è resi conto che quando il numero di esempi è troppo elevato, è necessario scartare parte dei risultati. Pertanto, pur avendo a disposizione opere intere, la maggior parte degli esempi in esse contenuti potrebbero non essere considerati per garantire un bilanciamento nei risultati. Ci è sembrato dunque più utile incrementare le dimensioni del CP aggiungendo più estratti di opere diverse, anziché opere in versione integrale.
5. *Corpus* di prosa letteraria e saggistica: i testi sono stati scelti in base a quanto tradotto dagli inizi del XX secolo in Russia e in Italia e all'impatto che hanno avuto alcune opere nella cultura di arrivo. Quest'ultimo dato si evince dal gran numero di versioni esistenti di una stessa opera (cfr. *Il Cappotto* di N.V. Gogol', tradotto in italiano più di trenta volte). Per garantire maggiore rappresentatività riguardo ai generi, abbiamo deciso di includere anche testi di saggistica appartenenti a diversi ambiti (arte, critica letteraria, epistolario, filosofia, legge / scienze sociali, memorie, religione / teologia, scienze applicate, scienze esatte e naturali, storia / biografia)[17].
6. *Corpus polivariantnyj*: visto l'elevato numero di versioni pubblicate in Italia per alcune opere russe, si è deciso che in una seconda fase il CP ita-ru sarà arricchito da più traduzioni di una stessa opera. Ciò è già avvenuto per il racconto di Gogol' *Il Cappotto* (15 versioni) e per *I racconti di Kolyma* di V. Šalamov (2 versioni).

In seguito al nostro lavoro da dicembre 2015 il CP ita-ru è salito a 4.066.172 parole[18], con 30 opere letterarie russe con la traduzione italiana (estratti e opere intere) e 21 estratti di opere letterarie italiane con la traduzione russa. La sezione di saggistica contiene al momento solo 5 estratti in lingua russa con traduzione italiana. Si prevede di inserire a breve sedici estratti in lingua italiana e undici in lingua russa (due opere per ogni genere in entrambe le direzioni).

---

[17] Per la scelta dei macro-settori ci siamo riferiti all'*Index Translationum*, un database dell'UNESCO che segnala i libri tradotti dal 1979 al 2012, dividendoli per generi.

[18] In linguistica dei *corpora* generalmente si intende per 'numero di parole' il numero di *token*, ossia di occorrenze effettive e non di lemmi (definiti *types*).

Tuttavia il CP non rispetta del tutto i criteri sopracitati per via dell'intervento di una collaboratrice esterna che lavorando indipendentemente ha inserito testi in un'unica direzione (dal russo all'italiano) e in versione integrale. In futuro si prevede di ovviare a questo parziale sbilanciamento inserendo un numero maggiore di opere italiane e creando un *sottocorpus* di testi paralleli in versione integrale.

3. *Applicazioni dei corpora paralleli*

In questo paragrafo ci concentreremo su alcune possibili applicazioni dei CP soprattutto, ma non solo, nell'ambito della traduzione e della linguistica contrastiva e tipologica. Il nostro intento è mostrare le potenzialità del CP sopra descritto, riferendo di ricerche ed esperimenti realizzati con questo e con altri CP.

Per quanto riguarda la traduzione, i *corpora* possono essere strumenti preziosi per identificare l'equivalente di una data parola in un contesto specifico o per utilizzare la lingua in modo naturale, tenendo conto di quali sono le forme di espressione più diffuse in un determinato ambito (Bowker, Pearson 2002: 193-197). In particolare, il CP si rivela utile nei casi in cui gli equivalenti standard forniti dal dizionario non sono accettabili in uno specifico contesto traduttivo. L'analisi dei dati del *corpus* consente infatti di identificare nuovi equivalenti e di vedere in quali contesti il loro uso risulta opportuno, riformulando così le condizioni di equivalenza (Dobrovol'skij *et al.* 2005: 272)[19].

Ciò favorisce l'utilizzo dei *corpora* anche nella didattica della traduzione. Ne riportiamo un esempio: nel corso di Traduzione editoriale per la Laurea Magistrale in *Specialized Translation* (Università di Bologna, sede di Forlì) si è lavorato sulla traduzione della congiunzione причем[20] nel seguente passaggio:

(1) Приведем лишь два примера роли Владимирской иконы в истории русской государственности, причем в каждом случае воздействие образа на публику осуществлялось во многом посредством его украшений.

Il dizionario monolingue di Ožegov fornisce la definizione: "То же, что притом", e per притом: "Вместе с тем, к тому же, в добавление к тому" (Ožegov, Švedova 1992); i dizionari bilingui fornivano poi diversi eteronimi: "oltre a ciò", "nello stesso tempo" (Majzel', Skvorcova 2001), a cui ABBY

[19] Per altri esempi di utilizzo dei CP negli studi di linguistica tipologica si veda Sičinava 2013.

[20] Dell'uso di questo tipo di connettori si sta occupando O. In'kova in "Otnošenie soputstvovanija i ego pokazateli v russkom, francuzskom i ital'janskom jazykach", VI Congresso Internazionale di Linguistica testuale contrastiva Lingue slave – Lingue romanze, 16-18 maggio 2019, Cracovia.

LINGVO x5 aggiunge un esempio con la soluzione “in più”: “не прав, причем еще спорит”, “(non solo) non ha ragione, ma in più discute pure”. Nel dizionario di Dobrovolskaja si trova invece un’unica traduzione, “anzi”: “он сдал экзамен, причем очень хорошо”, “ha superato l’esame, anzi, l’ha superato benissimo”. Data l’inadeguatezza delle soluzioni proposte per il contesto in esame, si è proceduto alla ricerca di причем sul CP italiano-russo selezionando dapprima il *sottocorpus* dei testi con l’italiano come lingua di partenza e traduzione russa. Sono state ottenute 20 occorrenze: in 6 casi причем risulta aggiunto rispetto al testo italiano, dove non ha corrispondenti, in altri 6 rende “e” del testo di partenza, in 2 occorrenze corrisponde a “ma”, in 2 a “che”, in 1 a “di cui”, in 1 ad “anzi” e in 2 casi a due espressioni complesse: una caratterizzata dalla ripetizione del verbo (2) e l’altra dalla struttura “non …, ma …” (3):

(2) Certo voleva qualcosa e lo voleva subito. [Umberto Eco, *Il nome della rosa* (1980)]
Безусловно, ему было что-то очень нужно, причем немедленно. (trad. russa di E. Kostjukovič, 1989)

(3) […] e continuava a mordersi le unghie (se le mangiava non alzando un dito alla bocca, ma abbassandolo, con la mano a rovescio, il gomito alzato). (I. Calvino, *Il barone rampante*, 1957)
[…] теперь она сердито грызла ногти, причем подносила палец ко рту сверху вниз, поднимая для этого руку и выставив локоть. [I. Kal’vino, *Baron na dereve*, 1965 (trad. it. di Lev Veršinin)]

Cercando poi nel *sottocorpus* dei testi russi con traduzione italiana si ricavano 128 occorrenze. Qui причем è reso con le stesse forme già individuate nel *sottocorpus* dei testi di partenza italiani (in particolare la congiunzione “e”, e l’assenza di corrispondenti), ma anche con espressioni come: “e oltre a ciò”, “e così dicendo” (dove il gerundio riprende il predicato precedente), “(e) per di più”, “tanto più che”, “anche se”, “durante i quali”, “mentre”, “e intanto”, “e allora”, “in modo che” ed altre.

Risulta così evidente che причем codifica un significato molto ampio di contemporaneità e/o aggiuntivo, il quale, se spesso risulta ridondante e non richiede la traduzione in italiano, in altri contesti può essere tradotto con “nello stesso tempo”, “e”, “e intanto”, “mentre”; “oltre a ciò”, “per di più”, “tanto più che”; con una relativa (“che”[21], “di cui”), con espressioni complesse come in (2) e (3) e, infine, per arricchimento inferenziale[22], può esprimere altri tipi di relazioni transfrastiche e aprirsi a rese quali “anche se” e “in modo che”. Nel nostro caso specifico, la soluzione più naturale è risultata quella in cui причем non viene tradotto:

---

[21] Nelle frasi relative appositive “il pronome relativo cumula la funzione di collegamento anaforico con l’antecedente a quella di congiunzione” (Prandi 2011).

[22] Cfr. Prandi (2004: 46-47).

(4) Di seguito sono riportati due esempi che evidenziano l'importanza del ruolo dell'icona di Vladimir nella storia dello stato russo. In entrambi i casi l'effetto prodotto dall'immagine sul pubblico derivava in gran parte dai suoi ornamenti.

In modo analogo il CP trova importanti applicazioni nella didattica del russo L2, poiché l'osservazione delle forme più complesse e oscure per i discenti — quali ad esempio i segnali discorsivi — in una grande quantità di contesti d'uso consente una più rapida comprensione del loro funzionamento in russo, grazie anche alla presenza del testo parallelo in italiano.

Il *corpus* fornisce anche utili spunti per l'analisi di diverse traduzioni dello stesso testo: nel CP italiano russo questo per ora è possibile nel caso de *Il cappotto* di N. Gogol' (15 versioni) e *I racconti di Kolyma* di V. Šalamov (2 versioni).

Un'altra applicazione del CP nell'ambito della traduzione riguarda lo studio delle caratteristiche specifiche dei testi tradotti. Secondo Teich (2003: 143-147) questi ultimi sarebbero caratterizzati da un maggior grado di semplificazione, con una minor varietà e densità lessicale; da normalizzazione, ossia la tendenza a rendere conformi alle caratteristiche della lingua di arrivo gli elementi più peculiari del testo di partenza; da un maggior grado di esplicitazione; dall'influsso delle strutture linguistiche della lingua di partenza su quelle del testo di arrivo. Questo tipo di studi prevede l'utilizzo di un *corpus* monolingue accanto a quello parallelo (alcuni esempi sulla base del *corpus* russo-inglese in Dobrovol'skij *et al*. 2005: 274-275).

Per quanto riguarda la linguistica tipologica e contrastiva, accanto ai metodi tradizionali (elicitazione di informazioni dai madrelingua o confronto tra descrizioni teoriche esistenti), sta prendendo piede l'analisi di tipo quantitativo basata sui dati dei *corpora* (Teich 2003: 10-12). Nell'ambito della lessicologia contrastiva i CP, associati ai grandi *corpora* monolingui, permettono di identificare i diversi significati di un lessema in base ai contesti in cui occorre, individuarne le collocazioni e analizzarne la struttura argomentale. Interessante l'esempio proposto da Dobrovol'skij *et al*. (2005: 276-277) riguardo alle collocazioni di *absolutely* e абсолютно[23]. Dell'uso di "assolutamente" e dei suoi equivalenti in russo si sta occupando V. Benigni in "Da intensificatore assoluto a segnale discorsivo: il caso di 'assolutamente!' e dei suoi equivalenti in russo", VI Congresso Internazionale di Linguistica testuale contrastiva Lingue slave – Lingue romanze, 16-18 maggio 2019, Cracovia. Il *British National Corpus* evidenzia tre contesti d'uso di *absolutely*: come intensificatore di un aggettivo che già esprime un grado elevato di una caratteristica (*appalling*, *delighted*), come intensificatore di espressioni di 'necessità' (*necessary*, *vital*) e come intensificatore di forme che esprimono 'comprensione', 'sicurezza' (*clear*, *sure*). Nel CP, invece, l'avverbio russo абсолютно mostra delle differenze nelle sue collocazioni: nel caso dei contesti di secondo tipo, a fronte di *absolutely* in traduzione, nei testi di partenza

---

[23] Dell'uso di "assolutamente" e dei suoi equivalenti in russo si sta occupando V. Benigni in "Da intensificatore assoluto a segnale discorsivo: il caso di 'assolutamente!' e dei suoi equivalenti in russo", VI Congresso Internazionale di Linguistica testuale contrastiva Lingue slave – Lingue romanze, 16-18 maggio 2019, Cracovia.

russi non si incontra l'avverbio абсолютно, ma altre forme (тебе непременно нужно) e con l'avverbio predicativo необходимо ricorrono solo i modificatori жизненно, чрезвычайно, принципиально. Per i contesti del terzo tipo nel CP è presente solo un esempio dove ad *absolutely* corrisponde прекрасно (*he understood absolutely*, оно прекрасно поняло).

Il CP si presta anche a ricerche in ambito morfo-sintattico, come illustrato da V.A. Plungjan[24] riguardo al valore di presente assunto dal futuro perfettivo russo preceduto dalla particella negativa *ne*. Oltre ai casi già descritti in letteratura, nei quali il perfettivo perde il suo valore futurale ("то сядет, то встанет" 'un po' sta seduto, un po' sta in piedi' e la costruzione negativa con significato di impossibilità), Plungjan, grazie all'uso dei CP, ha individuato, in altri contesti, tre tipi di forme alla prima persona, che egli definisce '*immediatnye*', poiché l'azione negata non avviene già nel momento dell'enunciazione: не скажу, не дам, не пущу. In questi contesti i verbi assumono un valore vicino a quello dei performativi:

(5) Qual è? – Non te lo dico. – Ora me lo devi dire. (N. Ammaniti, *Io non ho paura*, 2001)
И кто это? – Не скажу. – Почему? Так нечестно. [N. Ammaniti, *Ja ne bojus'*, 2005 (trad. it. di V. Nikolaev)]

Il verbo al presente nei testi italiani, francesi e inglesi, a fronte del futuro russo nei diversi CP, consente di evidenziare il valore non futurale di queste costruzioni russe – non descritte nelle grammatiche e non immediatamente identificabili dai madrelingua – e di fornirne un'accurata descrizione.

L'utilizzo dei CP si è rivelato molto proficuo anche nell'ambito della frase complessa e a livello testuale, in particolare per lo studio dell'espressione delle relazioni transfrastiche. In Biagini 2012, ad esempio, è stato evidenziato come la relazione finale in russo, oltre che da чтобы, possa essere codificata da congiunzioni polirematiche come для / ради / в благо того чтобы; da locuzioni congiuntive come с целью, с намерением, с желанием, с расчётом; da gerundi come планируя, желая, мечтая, боясь; e da пусть e пускай. Inoltre, la relazione finale può essere espressa da mezzi diversi dalla frase complessa, come la coordinazione e la giustapposizione, eventualmente arricchiti con espressioni anaforiche e termini predicativi appropriati (Prandi *et al*. 2005: 29-30). La pluralità di mezzi che il sistema offre al parlante per esprimere un determinato contenuto concettuale può anche coincidere in due lingue diverse; tuttavia l'opzione considerata più adeguata, ossia la più frequente, è spesso differente. La scelta è determinata dall'uso collettivo, dalla realizzazione concreta delle possibilità messe a disposizione dal sistema e filtrate dalla norma (cfr. anche Dobrovol'skij *et al*. 2005: 273-274). In Biagini 2012 grazie all'utilizzo di un piccolo CP è emerso come nell'espressione della relazione finale in italiano prevalgano le strutture ipotatti-

---

[24] Ciò è avvenuto nell'ambito della lezione *Parallel'nyj korpus v naučnych issledovanijach*, tenuta da V.A. Plungjan il 10 novembre 2016 presso l'Università di Bologna, Dipartimento di Interpretazione e Traduzione, Forlì.

che, mentre in russo quelle paratattiche. Lo stesso tipo di indagine è stato svolto per le relazioni concessiva e consecutiva (Biagini 2016a e 2016b). In questi studi la definizione del contenuto della relazione transfrastica viene utilizzata come *tertium comparationis* per individuare e confrontare la sua realizzazione formale nelle due lingue. Effettuando una ricerca dei mezzi di espressione prototipici (ad esempio "sebbene" per la relazione concessiva in italiano), si osservano le forme corrispondenti nei testi paralleli russi. Ciò consente di verificare quali strutture prevalgano nelle due lingue, mettendo in evidenza il 'genio della lingua', l'impronta lasciata sui testi dalle scelte dei parlanti. Sono state così identificate forme di espressione che tradizionalmente non sono descritte dalle grammatiche, come ad esempio "a furia di" come resa di "так ..., что" per la relazione consecutiva.

4. *Il* Corpus *parallelo italiano-russo per l'incremento e il consolidamento della ricerca*

Il progetto sopra esposto mira a creare un CP all'interno del NKRJa di almeno 20 milioni di parole e bilanciato dal punto di vista della tipologia testuale, la cui interrogazione possa produrre risultati statisticamente rilevanti. Per un simile scopo è auspicabile lo sforzo congiunto di chi già, fra gli slavisti italiani, usa i *corpora* per le proprie ricerche. Per promuovere una simile collaborazione sarebbe utile la creazione di un portale: oltre a raccogliere testi allineati alla loro traduzione e addestrare all'uso dei CP per la lingua russa già esistenti (mediante istruzioni, esempi e *tutorial*), contribuirebbe anche a un dibattito scientifico che, basandosi sui CP, diffonda e promuova nuovi spunti di ricerca (tramite Forum, *webinari*, la pubblicazione *open access* dei risultati delle ricerche contrastive russo-italiane).

Un simile portale potrebbe raccogliere e mettere in condivisione sia i *corpora* particolari creati dai singoli studiosi o dai gruppi di ricerca (per lo studio contrastivo del lessico, della traduzione, dei linguaggi speciali ecc.) sia eventuali *Sovracorpora database* (*nadkorpusnye bazy dannych*), ossia *corpora* speciali, creati sulla base di CP già esistenti ma con annotazioni appositamente studiate per la ricerca di aspetti linguistici particolari. Questi *corpora* di ultimissima generazione sono utili soprattutto per quegli elementi del sistema linguistico che sfuggono alle annotazioni morfosintattiche e semantiche dei CP, che spesso eccedono i confini di parola e sono trascurati dalle grammatiche tradizionali, come per esempio connettori, particelle, espressioni polirematiche ecc. Un esempio lo abbiamo nel *Sovracorpora database* per lo studio contrastivo dei connettori in francese e russo (Zaliznjak *et al*. 2015; Zacman *et al*. 2016).

Il portale, accanto agli studi linguistici e letterari di tipo contrastivo basati sui CP, svilupperebbe infine un secondo ambito, altrettanto importante per la russistica italiana: gli studi acquisizionali sulla base dei 1) *corpora* del russo come L2 prodotti da studenti italofoni, 2) *corpora* del russo degli *ėritažniki* (un fenomeno in grande espansione in Italia) e 3) *corpora* degli errori (Rachilina 2016).

Le applicazioni dei summenzionati filoni di ricerca sarebbero innumerevoli, dalla creazione di grammatiche e vocabolari contrastivi sulla base dei dati dei CP, agli strumenti per la traduzione e il sostegno tecnologico e didattico all'insegnamento del russo come lingua straniera.

Bibliografia

Baroni *et al.* 2009: M. Baroni, S. Bernardini, A. Ferraresi, E. Zanchetta, *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora*, "Language Resources and Evaluation", XLIII, 2009, 3, pp. 209-226.

Biagini 2012: F. Biagini, *L'espressione della finalità in russo. Uno studio contrastivo con l'italiano*, Bologna 2012.

Biagini 2016a: F. Biagini, *L'espressione della relazione concessiva fattuale in italiano e in russo*, in: V. Benigni, L. Gebert, Ju. Nikolaeva (a cura di), *Le lingue slave tra struttura e uso*, Firenze 2016, pp. 45-62.

Biagini 2016b: F. Biagini, *L'espressione della relazione consecutiva in un corpus italiano-russo di testi per l'infanzia*, in: O. Inkova, A. Trovesi (a cura di), *Langues slaves en contraste, Славянские языки in comparatione, Lingue slave a confronto*, Bergamo 2016, pp. 243-264.

Bowker, Pearson 2002: L. Bowker, J. Pearson, *Working with Specialized Language. A practical guide to using corpora*, London - New York 2002.

Dobrovolskaja 1997: Ju. Dobrovolskaja, *Dizionario russo-italiano*, Milano 1997.

Dobrovol'skij 2009: D.O. Dobrovol'skij, *Korpus parallel'nych tekstov v issledovanii kul'turno-specifičnoj leksiki*, in: *Nacional'nyj korpus russkogo jazyka: 2006-2008. Novye rezul'taty i perspektivy*, Sankt-Peterburg 2009, pp. 383-401.

Dobrovol'skij *et al*. 2005: D.O. Dobrovol'skij, A.A. Kretov, S.A. Šarov, *Korpus parallel'nych tekstov: architektura i vozmožnosti ispol'zovanija*, in: *Nacional'nyj korpus russkogo jazyka: 2003-2005*, Moskva 2005, pp. 263-296.

Dobrovol'skij, Levontina 2009: D.O. Dobrovol'skij, I.B. Levontina, *Russkoe* net, *nemeckoe* nein, *angliskoe* no: *sopostavitel'noe issledovanie semantiki na baze parallel'nych korpusov*, in: *Komp'juternaja lingvistika i intellektual'nye technologii: po materialam ežegodnoj Meždunarodnoj konferencii* «*Dialog 2009*», 8, Moskva 2009, pp. 97-101.

Dobrovol'skij, Levontina 2015: D.O. Dobrovol'skij, I.B. Levontina, *Modal'nye* časticy *i ideja aktualizacii zabytogo (na materiale parallel'nich korpusov)*, in: *Komp'juternaja lingvistika i intellektual'nye technologii: po materialam ežegodnoj Meždunarodnoj konferencii* «*Dialog*», 14, I, Moskva 2015, pp. 104-117.

Johansson 1998: S. Johansson, *On the role of corpora in cross-linguistic research*, in: S. Johansson, S. Oksefjell (eds.), *Corpora and Cross-linguistic Research: Theory, Method and Case Studies*, Amsterdam-Atlanta GA, 1998, pp. 3-24.

Majzel', Skvorcova 2001: B. Majzel', N.A. Skvorcova, *Bol'šoj russko-ital'janskij slovar'*, I-II, Moskva 2001.

McEnery, Wilson 2001: T. McEnery, A. Wilson, *Corpus Linguistics*, Edinburgh 2001.

Noseda 2016: V. Noseda, *Corpora paralleli e linguistica contrastiva: ampliamento e applicazioni del corpus italiano-russo nel Nacional'nyj Korpus Russkogo Jazyka*, Milano 2016.

Ožegov, Švedova 1992: S.I. Ožegov, N.Ju. Švedova, *Tolkovyj slovar' russkogo jazyka*, Moskva 1992.

Prandi 2004: M. Prandi, *The Building Blocks of Meaning*, Amsterdam-Philadelphia 2004.

Prandi 2011: M. Prandi, *Subordinate, frasi*, *Enciclopedia dell'italiano*, <http://www.treccani.it/enciclopedia/frasi-subordinate_%28Enciclopedia_dell'Italiano%29/> (ultimo accesso: 15.03.19).

Prandi *et al*. 2005 M. Prandi, G. Gross, C. De Santis, *La finalità. Strutture concettuali e forme d'espressione in italiano*, Firenze 2005.

Rachilina 2016: E.V. Rachilina, *O novych instrumentach opisanija russkoj grammatiki: korpus ošibok*, "Russkij jazyk za rubežom", 2016, 3, pp. 20-25.

Reznikova 2009: T.I. Reznikova, *Slavjanskaja korpusnaja lingvistika: sovremennoe sostojanie resursov*, in: *Nacional'nyj korpus russkogo jazyka: 2006-2008. Novye rezul'taty i perspektivy*, Sankt-Peterburg 2009, pp. 402-461.

Sičinava 2005: D.V. Sičinava, *Nacional'nyj korpus russkogo jazyka: očerk predystorii*, in: *Nacional'nyj korpus russkogo jazyka: 2003-2005. Rezul'taty i perspektivy*, Moskva 2005, pp. 21-30.

Sičinava 2013: D.V. Sičinava, *Parallel'nye korpusa Nacional'nogo korpusa russkogo jazyka kak instrument leksičeskoj tipologii*, in: R. Guzmán Tirado, I.A. Votyakova (eds.), *Tipologìa léxica*, Granada 2013, pp. 77-82.

Teich 2003: E. Teich, *Cross-Linguistic Variation in System and Text*, Berlin - New York 2003.

Zacman *et al*. 2016: I.M. Zacman, O.Ju. In'kova, M.G. Kružkov, N.A. Popkova, *Predstavlenie krossjazykovych znanij* o *konnektorach v nadkorpusnych bazach dannych*, "Informatika i eë primenenija", I, 2016, 10, pp. 106-118.

Zaliznjak *et al*. 2015: A.A. Zaliznjak, I.M. Zacman, O.Ju. In'kova, M.G. Kružkov, *Nadkorpusnye bazy dannych kak lingvističeskij resurs*, in: *Korpusnaja Lingvistika-2015*, Sankt-Peterburg 2015, pp. 211-218.

Abstract

Francesca Biagini, Anna Bonola, Valentina Noseda
The Italian-Russian parallel corpus in the national Russian corpus: a project for its enlargement, applications and future development

In this paper, we will introduce our project for enlarging the Italian-Russian parallel corpus of the NKRJA, presenting as well some current or possible and future applications. The enlargement, carried out since 2014, has allowed us to abandon the pilot phase and to reach about 4 million words. However, we expect to reach 20 million, which is the required size to have a reliable and balanced corpus, another essential qualitative feature of parallel corpora. Illustrating its possible applications, we will present some experimental uses for teaching translation at the University of Bologna and some research in the sphere of Russian-Italian Contrastive Linguistics (about the non-temporal value of Russian future tense and the expression of the relationship between clauses). Finally, we will briefly present a project based on the creation of a portal in order to constantly enlarge the parallel corpus and promote its use in linguistic and literary research, in Translation Studies and in the teaching of Russian as a second language.